Integrated Information Infrastructure Services

Background and Rationale

As emphasized in the Administration's "Agenda for Action" white paper, perhaps the most significant technological shift in this decade will be in the way people deal with information, from the personal level, business level, and the government level. Information has become one of the nation's most critical economic resources, for service industries as well as manufacturing, for economic as well as national security. In an era of global markets and global competition, the technologies to create, manipulate, manage and use information are of strategic importance for the United States. Those technologies will help U.S. businesses remain competitive and create challenging, high- paying jobs. They also will fuel economic growth which, in turn, will generate a steadily-increasing standard of living for all Americans. For all these same reasons, these technologies will also be of strategic importance in maintaining the Laboratory's ability to be both stewards and providers of advanced weapon technologies.

Technologically, the NII consists of basically of three general components: hardware infrastructure (primarily the network), software infrastructure (primarily software services), and the applications (working interoperably and over the network). The NII will consists of the integrated working together of these components for everyone's benefit. "*Every component of the information infrastructure must be developed and integrated if America is to capture the promise of the Information Age*"

The administration is committed to working with private industry to establish the basic principles of the NII, and to use regulation to encourage private sector investment, technological innovation, to ensure universal service, to encourage interoperability, and to ensure information security and trustworthiness.

The strategy we are choosing to meet this technological challenge is to develop an integrated approach to applying the information technology to science and society. This includes using requirements from a suite of applications to define a common infrastructure. We then will develop an array of technologies to meet these requirements. This process allows, for example, a solution of problem of importance to health-care to rapidly impact problems in industrial manufacturing, or even gas and oil exploration and production. This involves not only finding common solutions but in tightly coupling the research and development and information management so that these new tools can be more rapidly deployed both in the laboratory and by industry, itself. More information on this work can be obtained from the World Wide Web in the URL: "`http://www.acl.lanl.gov/sunrise/sunrise.html`".

Program development opportunity

The four components of the Federal NII program include

- Information Access to improve public access to government information by allowing agencies to make such information available electronically ,
- National Performance Review to use information technology investments to move towards *Electronic Government*,
- High Performance Computing and Communications Program,
- and Service Delivery to use information technology to deliver services such as loans, benefits, and grants directly to the public.

The Federal High Performance Computing and Communications program has allowed DOE to continue its leadership in this core area, long supported by the weapons program. This has resulted in a flag ship effort in the national high performance computing arena at Los Alamos. Because DOE is getting no additional funds in FY94 or FY95 for NII activities, the DOE's leadership position is eroding rapidly. In FY95 the President's budget contains over $3 billion for NII activity which we are aggressively pursuing. This includes requesting the inclusion of $50M into the federal budget for the DOE in FY95. The linkage of these various components is crucial to the economic competitiveness of the nation and to the actual use of high-performance computing in its various forms throughout industry. As is clear from the emphasis of the national program, the ability to deal with large amounts of distributed information intelligently and efficiently will be fundamental to not only the success of High Performance Computing but to the societal infrastructure as a whole later in this decade. The potential revenue to the laboratory in this area alone is well in excess of $10 million/year and can leverage at least 10 times that in related efforts in programs that use the infrastructure tools.

By developing an integrated approach to the Information Infrastructure and linking High Performance Computing with commodity software and hardware tools, we have an opportunity not only for Federal funding but also working directly with industry to assist them in their large data management and a analysis problems. This is possible because of our unique position in understanding the computational algorithms and integration technology.

Connection to strategic directions and laboratory goals

The Laboratory has a goal to sustain its world-class strength in High Performance Computing and to work closely with industry on a variety of areas including materials manufacturing and a variety of others. We have identified several industrial partners which are working with us to advance the state of the art in their field. These include manufacturing companies like Gatan that market Tunneling Electron Microscope equipment, to Xerox who is seeking new ways of developing Xerographic engines, to Adobe and HaL who are developing advanced tools for electronic publishing and network navigation, to the National Jewish Center for Immunology and Respiratory Medicine, and to the Dept of Commerce.

R&D approach and likelihood of project success

The HIIII project was started in FY94 to enable the laboratory to take an aggressive stance in NII and provide a strong R&D base for obtaining NII funds from various government agencies and even from U.S. industry.

In order to achieve this ,we have set the following three objectives:

- To develop common information-enabling tools for advanced scientific research and its application to industry
- To enhance the capabilities of important research programs at the Laboratory
- To define a new way of collaboration between computer science and scientific research and development.

The basic paradigm being developed involves a document-centric user interface which will allow arbitrary object support including embedded applications, multimedia video/voice fragments and links to a wide information space.

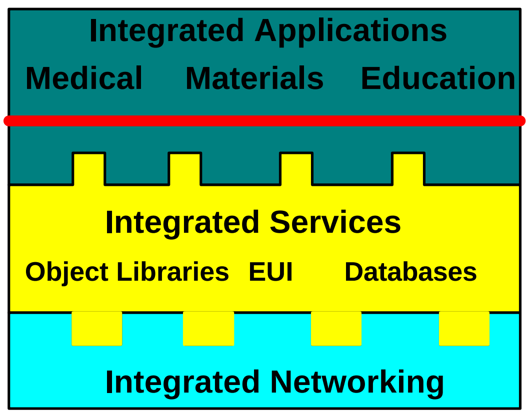
We are developing an "information kiosk or portal" based on an ATM network so that all the participants can exchange, "publish", or interact on applications and data. This will function on heterogeneous platforms, provide for constrained access to data through security mechanisms, and be extensible. The data-mining technology will include the ability to quickly browse large complex image databases with various feature extraction capabilities, provide advanced, selective, compression algorithms, and the ability to merge and purge large complex data sets.

The focus is on real scale problems, reflecting our belief that only experience with real problems encompassing huge data sets will facilitate true progress. Dealing exclusively with toy problems is as likely to mislead as it is to inform. For example, trivial issues like name fusion, can become critical with huge data sets. Therefore, our focus will be on developing prototype tools that can be tested in real-sized testbeds.

We believe the benefits will be to the laboratory as a whole in helping define the way research will be done in the future, in enhancing the competitiveness of laboratory research programs, and in enhancing the nation's ability to use advanced information technologies for applications of importance to industry.

We are taking an integrated approach in which several applications in disparate areas are developed simultaneously. This is to avoid vertical solutions applicable to only one domain, but rather to raise the level of commonality and interoperability to a new level. The identified application areas are at the core of the NII and include telemedicine with the potential for substantial change in the way health care is carried out, materials modeling and analysis which can change the way materials data is analyzed and dispersed in the manufacturing community, transportation [analysis] which has the potential to change the way traffic information is utilized and understood to improve the functioning of the cities, and finally education in which what is being taught as well as the methods of teaching are currently under rapid change. These four areas represent a sufficient diversity of applications to characterize the NII and to provide an excellent foundation for determining the common infrastructure required.

The integration of applications and their dependency on integrated services and networking is shown below.



We envision the use of the Sunrise environment by novice and expert users who wish to explore the NII and/or collaborate in research. The environment will provide the users with an interrelated set of tools which will enhance their interaction with a domain of knowledge. The interaction of the tools and the knowledge domains might produce a medical diagnostic system, an environment for materials development and design, or a classroom learning situation. The tools will facilitate the formulation of a problem, it's solution, the interpretation of results and the publishing of those results. In support of these tasks, the tools will provide mechanisms to query the range of information sources on the NII, to consult with experts in a knowledge domain, and to collaborate with co-workers in formulating and solving problems. The envisioned toolset has a number of components:

- A journalling facility, similar to a scientist's log book, to track the individual's research progress including the data sets accessed, the results of data analysis, the individual's interpretation of those results, and monitoring the individual's interactions with others. [Note that there are privacy implications here]. The log would be secure, and would provide authenticated time stamps which might be used to validate rights of first discovery of some idea. It will utilize object encapsulation to deal with multiple problems, patients, categories, etc. Appointments, meetings, etc. would be accommodated.

- An interface to experimental equipment in order to initiate, monitor and analyze an experiment connected over the network to the system. Similar functionality for initiating detailed computational simulations on remote supercomputers will also be provided.
- A remote data access and display capability which interacts with various distributed databases included OODB's over the network. It will provide a complex querying interface supporting text, multi-spectral images, and digital video attributes (perhaps combining with the results of previous queries in a hierarchical manner). It provides for the publishing of data by the user and the ability to "purchase" remote data either by value or by reference.
- A digital document repository which facilitates the retrieval and publishing of research documents. This might be just a variation on the more general data access tool. These documents might be hypermedia tours of selected parts of the individual's logbook.
- An accountant which will monitor the cost or potential cost of actions and invoke an authorization process under certain conditions and provide rapid feedback on the actual cost of the actions. (This is a variant of a work-flow capability) Such an authorization might be triggered when a remote data request accesses commercially billed information, which might only be advertised by attributes. The actual data in that case might be the part that costs.
- A toolbox of analysis modules which provides processing of data to facilitate the actual research.
- A facility that can provide a range of help from concise command summaries all the way to live telecollaboration with paid consultants
- User annotations across the entire domain would be supported. These also could be made available to collaborators or published more widely. We envision native viewers of much of the data with more sophisticated domain-specific analysis tools being invoked as needed.

More information on our approach to providing an integrated information infrastructure can be obtained through the World Wide Web in the URL: "`http://www.acl.lanl.gov/sunrise/sunrise.html` "

There are four components to the integrated infrastructure services: Distributed Computing, User Interface and Telecollaboration, Data Analysis and Visualization, and Security. They will all use a commonobject base and are described below.

Distributed Computing

David Forslund, ACL; Bob Tomlinson, C-8; John Reynders, ACL - $550K

The usefulness of the NII will be greatly enhanced when users of applications can quickly and easily locate and access data and computing resources in the network. Those resources may be, for example, public or private electronic libraries, data conversion services, or various other application specific services. Since thousands of services built by hundreds of organizations will exist on the network it is important to facilitate ways of building applications by using software components built by different organizations.

We must build an infrastructure that allows systems to be built and modified with interchangeable components in much the same way in which one combines stereo components when building an audio system. When building an audio system we can interchangeably add or subtract components without redesigning the system or even replace a tape player with a CD player or some other device without changing anything else in the system.

Users accessing the NII will be combining desktop machines such as PCs and workstations with various data and computational servers into a widely distributed yet integrated system. Components of the system will often have different CPU architectures, different data communication infrastructures and data rate capabilities, different operating systems, and be widely distributed. Nevertheless, users' applications will need to be able to call on those resources as if the resources were locally available.

This composable component based system requires further evolution of distributed computing environments. To accomplish this evolution, we are basing the Sunrise environment on the notion of distributed objects. Properly designed object based systems have the unique advantage that changes in one area of the system do not affect other areas. A properly designed set of object classes will enable us to link resources in a modular way to manage complexity and yet allow great extensibility. We are building the Sunrise distributed computing infrastructure on the emerging industry standard Common Object Request Broker Architecture.

Early emphasis in the development of objects has been on the Sunrise teleradiology project with the National Jewish Hospital in Denver. The system will allow CAT scan and X-Ray images of lungs to be submitted from anywhere in the network to an electronic repository. The image will then be processed to compute a "signature" that can then be used to find other similar images.

Current Status

We are building upon a base CORBA system, the IONA product Orbix, that will run on the computers of primary interest to Sunrise participants (Sun workstations, Silicon Graphics workstations, and IBM PCs).

We have worked through some serious problems interconnecting the main Sunrise graphical user interface system (GainMomentum) with external libraries including the Orbix library.

We have extended and improved an object system designed originally for Particle in Cell (PIC) simulation to work with a wider class of applications and systems. It has formed the initial basis for objects for the teleradiology part of the Information Applications project. By using appropriate layering of the software, we have been able to abstract away the machine architecture so that the same application can run on the CM-5, a workstation cluster, or a Cray Y/MP. Also a general IO library for storing the data from different machines in an architecture independent format has also been completed. We can store data to a local disk or to HPDS essentially invisible to the application developer.

Classes on Object Oriented Analysis and Design and Object Oriented Design in C++ were presented to Sunrise developers in order to accelerate the process of designing objects for other Sunrise applications and to arrive at a common methodology so that applications as well as various object components can interoperate.

Remaining Tasks for FY94

- Continue analysis, design, and implementation of objects for other Sunrise applications.
- Improve efficiency of general object system and extend based on evolving requirements
- Further analysis and design of objects of general usefulness for other Sunrise applications.

Tasks for FY95

- Extending objects for a wider variety of applications including multi-media objects.
- Demonstrate ORB interoperability and efficient execution over ATM networks
- Integration of object oriented database technology for storage of distributed objects.

User Interface and Telecollaboration

R. L. Phillips, C-5 - $50K

OVERVIEW

The overall goal of this subcategory is to develop a general, media-rich user interface system which has capabilities that meet the needs of all Sunrise application clients. In general, this will be an *executive user interface*, one that has a set of convenience capabilities common to all application areas, while still providing application clients easy access to their preferred user interfaces. One of the most important convenience capabilities is to facilitate telecollaboration between users. In its simplest form this feature will permit peer-to-peer collaboration in a common screen space. More advanced capabilities will include many-to-many collaboration and video teleconferencing. The Sybase product Gain*Momentum* is being used for interface development.

CURRENT STATUS

The executive user interface (EUI) has been designed and partially implemented. Among its features are:

- Compound (media-rich) document executive; provides a complete multimedia authoring and display capability. Data stored in Bento (the OpenDoc persistent storage system) "documents" can be dynamically assembled and viewed.
- A logbook facility, which draws upon the capabilities described above, will allow a researcher to maintain a time-stamped multimedia record of all activities. Since the logbook will be retained as a Bento file, it can be accessed by other applications
- Generic Sunrise data navigator; provides a search and retrieval tool, including mixed media searching capability. An appropriate viewer is invoked depending on media type.
- Customized, application-specific user interfaces can be defined and easily incorporated. Launch
- buttons can be added as needed to the basic EUI.
- Other EUI convenience features are a MIME-compliant mailer and viewer, multimedia annotations; post-its, sketchers, audio, video, animation, etc. For intra- and inter-document navigation, thumbnails and hyperlinks will be provided.

Some examples of the appearance of an EUI and some application-specific interfaces are shown in the following figures. Figure 1 shows an EUI and the possible contents of a logbook. Here the researcher elected to log an animated sequence of images representing temporal views of a global ocean simulation. A post-it note was added to the page by dragging an instance of the note icon from the left of the tools bar.

.

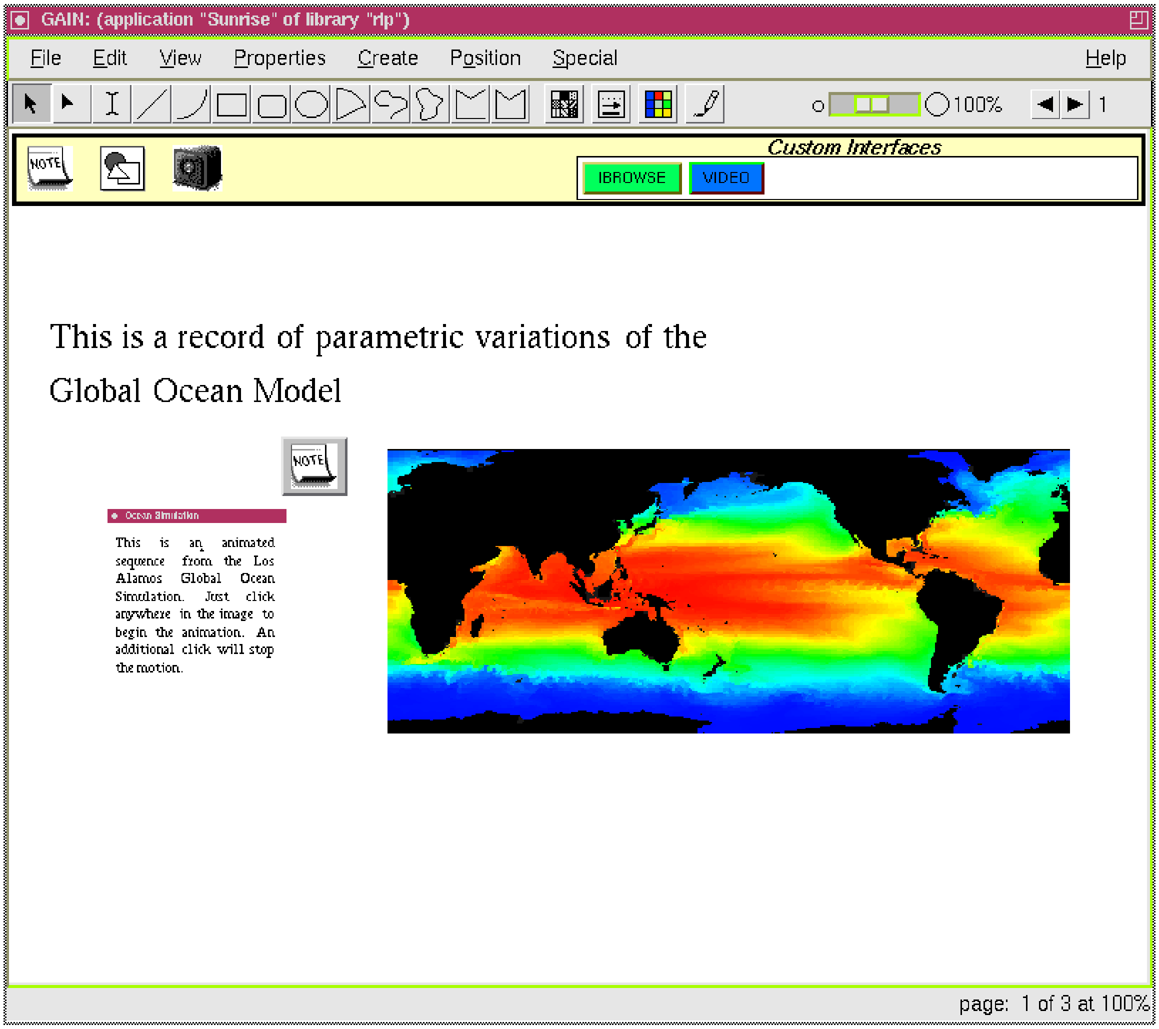


Figure 1.

The post-it is shown in its open state to reveal its contents. The buttons labeled **IBROWSE** and **VIDEO** select user interfaces for two of the Sunrise application clients. Their appearance is shown in Figures 2 and 3. Figure 2 shows the user interface that will be used by LANL researchers and physicians at the National Jewish Center in Denver for a teleradiology project. This is the interface to an image browsing application, which allows physicians to discover similarities in radiographic images.

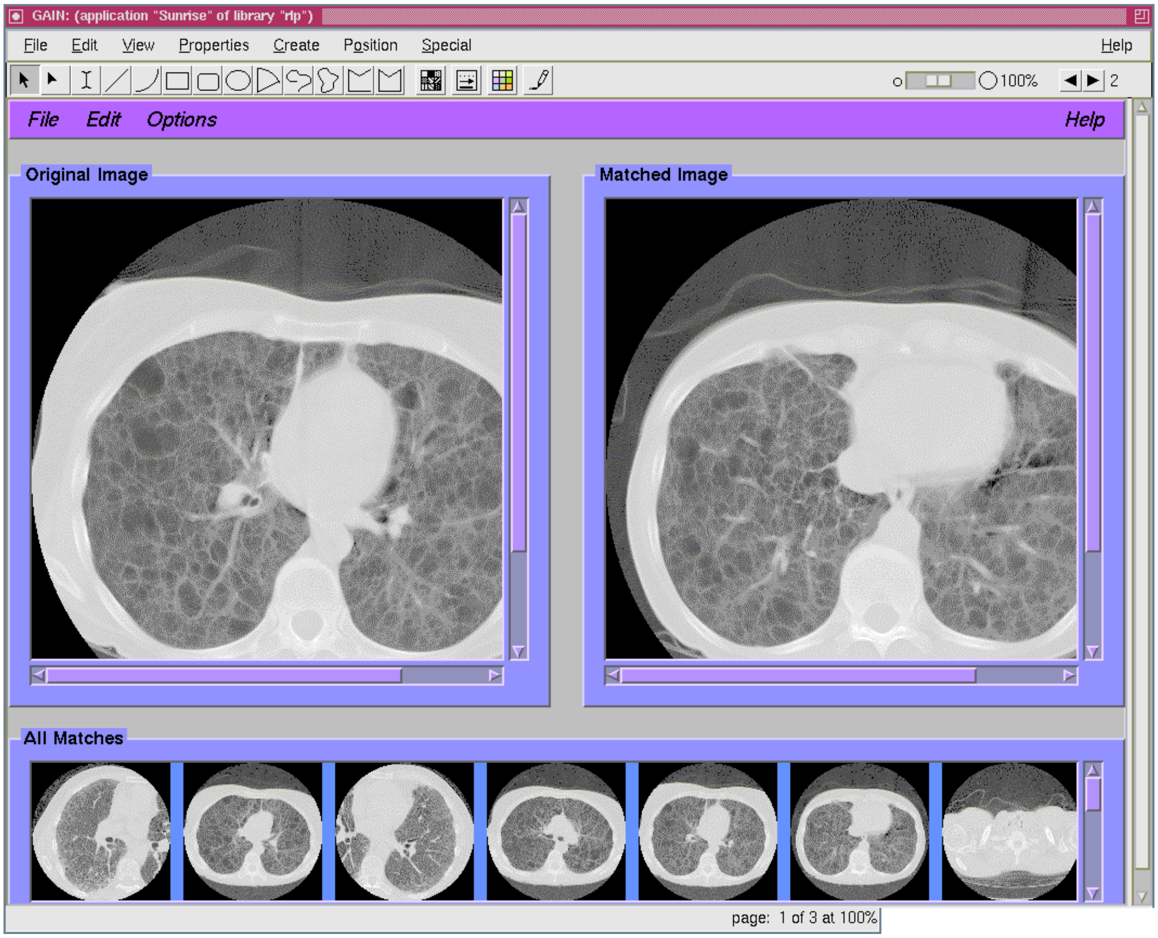


Figure 2.

Figure 3 is a mockup of an interface for accessing data from the LANL video server. The user selects the name of a video fragment from a scrolling list ("Ocean" has been selected in the example) and the sequence plays in a subwindow

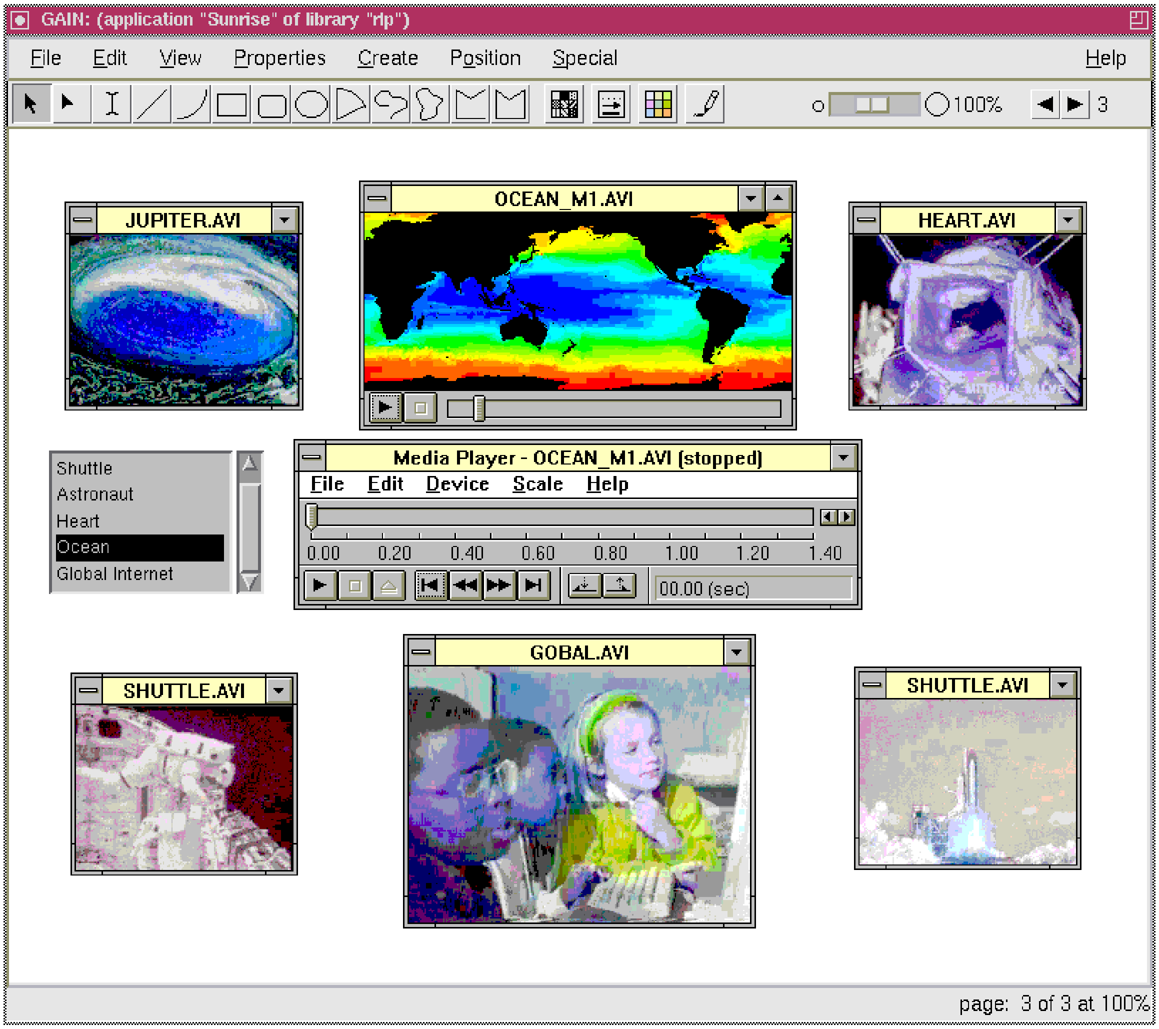


Figure 3.

REMAINING TASKS FOR FY 1994

Complete integration of application-specific interfaces into the EUI.

Implement a Bento storage/retrieval interface. This will require the development of dynamically linkable module for Gain*Momentum*

Complete the logbook facility.

Complete post-it, audio, video, animation, and sketcher convenience tools.

TASKS FOR FY 1995

During this fiscal year the bulk of activity will be devoted to implementation of telecollaboration tools. The first of these will facilitate peer-to-peer collaboration in a common screen space. Additional tasks will implement navigational tools and additional convenience features. Specifically, the following subtasks will be addressed:

- Development of additional application-specific interfaces to be incorporated into the EUI.
- Implementation of a data navigator facility, which will provide a search and retrieval tool, including mixed media search capability. This facility will be network-wide and will make use of CORBA browsing facilities.

- Implementation of thumbnails and hyperlinks for intra- and inter-document navigation.
- Implementation of a MIME-compliant mailer and viewer.
- Implementation of an HTML viewer for Mosaic compatibility. This will be based on source code
- that is available from NCSA for X Mosaic.
- Design and development of a remote procedure call library as infrastructure for telecollaboration
- capabilities.
- Development of a shared-workspace peer-to-peer telecollaboration capability.

Data Analysis and Visualization: *Active Visualization*

Al McPherson, C-3; Chuck Hansen, C-DO/ACL - $100K

Visualization plays a pervasive role in the NII; not only is it crucial for perusal of extremely large scientific data sets but visualization will impact the distributed computing, multimedia, data management, and networking efforts of NII. As a fundamental tool in the analysis of data for the selected focus projects, visualization is at the core of the proposed research. In addition to providing the necessary analytical tools, collaborative visualization tools will need to be developed which take advantage of the infrastructure and leverage it into a more powerful yet usable commodity.

Visualization spans the technology thrust areas and the application areas. Scientific visualization will build upon the efforts in distributed computing by working closely with the underlying foundations provided by the distributed object environment. Data management requires strong visualization both in terms of data perusal as well as in effective utilization of remote data. As mentioned, visualization will be integrated with multimedia. This will occur both in animation objects as well as active geometric objects which scientists can employ to effectively communicate underlying meaning within their data. Finally, ATM provides a media particularly suited to scientific visualization. The small cell size allows for dynamic data in the conventional sense of packet-switched digital video as well as providing a means for achieving temporal coherence with geometric entities crucial to the visualization process.

In addition to the technology base, scientific visualization plays an integral role with the focus applications. The data space and image space decomposition methods developed as part of HPCC will be domain-enhanced for this problem. The diverse sampling equipment employed by the materials modeling application will find enormous benefit from the integrated visualization environment which is being prototyped within the HPCC environment at the ACL. It is this prototype environment which will provide the basis for the remote NII prototype being proposed.

Active visualization allows a user to view the visual results of a simulation interactively. Active documents support the notion of interaction with their compound parts. Rather than view a static image, the reader can control aspects of the visualizations such as scaling, translation and rotation. The user may also be able to modify parameters of the visualization. These may be as simple as changing parameters to a plot while the graph representation changes dynamically. More advanced applications may allow the user to change visual parameters such as the opacity and isosurface values of volumetric renderings.

These parameters can be constrained to specific ranges by the document developer. Active visualizations may run locally or may be used to control a remote visualization with the resulting image displayed locally.

Integrate into the Sunrise environment:

- within compound document user interface
- using high speed local graphics hardware
- using high speed ATM networks to control remote rendering and for transfer of data
- using standard Sunrise tools for distributed computing data storage, user interface, etc.

Phases of Implementation

- Integrate OO design into Sunrise OO framework
- Native (workstation) rendering integrated into the Sunrise compound document interface
- Distributed rendering integrated with Sunrise network and distributed computing tools
- Combination of native (for interaction) and remote (CM-5 or T3D) rendering

Phases I and II are part of FY94 effort and are described below. Phases III and IV are proposed here as FY95 work. Subsequent phases may be defined and proposed as follow-on work to the FY95 proposal. We briefly describe each of the phases below. Tasks and deliverables are discussed in the sections.

Phase I

A goal of the Sunrise project is to produce an **integrated** set of tools for the construction of NII based scientific applications. To achieve this goal if integration, visualization tools must mesh seamlessly with other base technologies such as user interface, distributed computing, and data management. Phase I concentrates on design integration with other Sunrise components. For example, certain objects being designed may require render and display methods. Involvement throughout the design process will ensure an integrated approach to visualization.

Phase II

Sunrise software will make extensive use of a document-centric user interface. This interface enables the integration of many different components within an electronic document under a common user interface. We will integrate the visualization renderer into this interface, following user interface conventions established by the project whenever possible.

FY94 Efforts

As stated above, phases I and II are currently in work as part of the Sunrise FY94 effort. We are continuing to integrate the visualization design into the overall Sunrise object-oriented architecture. We have developed a simple native rendering capability using the Inventor toolkit on Silicon Graphics platforms. We are also working on integrating the native renderer into the Sunrise document-centric user interface.

Phase I is an ongoing task, evolving as the Sunrise design evolves. We expect to complete the integration of native rendering into the document-centric interface by July 1994 and demonstrate it with a small sample application by August 1994.

FY95 Tasks and Deliverables

Phases III and IV consist of new tasks proposed for FY95.

Phase 3

Tasks

- Link renderer and user interface with CORBA distributed object software.
- Link render with Sunrise data management

Deliverables

January 95

- Integrated native renderer and CORBA object access tools. Ability to extract Sunrise data from a repository and convert to local format. Ability to render locally. Schedule depends on progress of Sunrise data management design and construction.

March 95

- Demonstration application using native renderer and integrated Sunrise distributed computing and data management tools.

Phase 4

Tasks

- Modify native renderer to control remote visualization tasks on large machines (CM-5 or T3D). Use native renderer and user interface to interactively control remote rendering codes. Display results on local machine.
- Modify user interface to accept constraints on renderer parameters. Enhance data/render object to ease construction of parameter driven remote rendering.

Deliverables

July 95

- Completed data/render object supporting fully integrated remote rendering.

October 95

- Demonstration application using remote visualization interactively controlled by local render interface.

Security

Mike Neuman and Gary Christoph, C-1 - $150K

Goals:

It is our goal to provide the level of authentication, authorization, and privacy necessary for a large scale network providing the types of data exchange that are proposed. This security will enable other applications by offering new features such as digital patents, and by ensuring the privacy and integrity necessary for any sensitive information transfers such as medical records.

Implementation:

There are two divisions of implementation. The first is to build, from the ground up, a system for authentication, authorization, and privacy. These areas have not been dealt with from the standpoint of a universal network before, and consequently there is no previous work to build upon. Therefore, they must be implemented and viewed as highly experimental. The first division consists of:

- Implementing a secure, reliable system for distributing public keys
- Creating a means for universal authentication
- Creating a reasonable authorization and authentication structure
- Research into means of insuring user accountability

Once these have been finished, applications will be built on top of them.

- An e-mail system which automatically protects messages given integrity and privacy levels.
- Low level network routines for providing a simple programmer interface to integrity and privacy routines
- A universal, secure time server providing irrefutable time stamps on data

The second division is to implement tools designed to provide increased security to a network such as the internet which are then scalable to the NII. These are security testing and misuse detection tools such as:

- A host based security scanner for testing the operating system configuration and security feature implementation on a machine
- A network based security scanner for testing the vulnerability of a machine from a network standpoint
- Host-based auditing tools for automatically detecting host-based misuse on a fine grain level
- Network on-line auditing tools for automatically detecting network-based misuse in real-time
- Network off-line auditing tools for automatically detecting network-based misuse and host-based abuse on a coarse grain level.

FY94 efforts:

An initial foundation of authentication, integrity, and privacy tools are currently being developed. When completed, these tools can be used to provide these features to any number of applications using a simple API.

In addition to software development, means for providing authentication to large user populations at a low cost and high accuracy are being researched.

FY95 Tasks and Deliverables:

- Network level routines for providing integrity and privacy
- A preliminary secure, universal time server
- A host based security scanner for testing the operating system configuration and security feature implementation
- A network based security scanner for testing the vulnerability of a machine from a network standpoint
- Preliminary work on Network and host auditing tools

Related Laboratory capabilities and investigator's expertise

There are a number of information technology projects that have been started at the Laboratory during the past year. These include the Information Architecture project, the Info2000 project, the Library without Walls project, the DoE Information Infrastructure Initiative. Each has a different emphasis and in each case we have participants of the Sunrise project who are involved in or familiar with each of these. Although the projects do not overlap, we are able to share and exchange information on a regular basis. Another project started this past year is DoE's Gas and Oil National Information Infrastructure project. We have found that many of the technologies being developed in the Sunrise project are immediately applicable to this project. The variety of investigators have considerable expertise in Information Infrastructure technology and its use in technical applications. The base object technology being used in this project relates to a number of other object systems being designed around the laboratory. The investigator's in this project have more than 5 years experience in developing object systems. Nevertheless, we will continue to relate this work with the other work going on in at least three other divisions so that we will maximize both the impact of this work and minimize what needs to be developed. The data-mining capabilities provided, by C-3, for example, are unmatched anywhere in the country. Dick Phillips, a world-renowned multimedia document expert, is providing the executive user interface technology to these applications.